2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Sanger's Type Dynamical Systems for Canonical Variate Analysis

Mohammed A. Hasan† & Jawad A. K. Hasan‡

†Department of Electrical & Computer Engineering, University of Minnesota Duluth

‡Department of Electrical Engineering, University of Baghdad, Baghdad, IRAQ

E.mail:mhasan@d.umn.edu

Abstract

In this paper, several dynamical systems for computing canonical correlations and canonical variates are proposed. These systems are shown to converge to the actual components rather than to a subspace spanned by these components. Using Liapunov stability theory, qualitative properties of the proposed systems are analyzed in detail including the limit of solutions as time approaches infinity.

Keywords: canonical correlation analysis, polynomial dynamical systems, asymptotic stability, global stability, global convergence, invariant set, Lyapunov stability, Lasalle invariance principle

1 Introduction

Canonical correlation analysis (CCA) is a statistical technique that has been used in situations where a large number of variables of distinct types are to be investigated simultaneously. For example, in biomedical signal processing, a large number of physical and psychological variables are analyzed. Discriminant analysis and regression analysis are particular cases of this general technique. Thus, there is an interest in computing the most significant canonical correlations and corresponding variates of statistical data. Some of the computational approaches involve generalizing the singular value decomposition (SVD); see, Ewerbring and Luk [1].

Statistically speaking, CCA involves partitioning a collection of variables into two sets, an X-set and a Y-set. The objective is then to find linear combinations $a = x^T X$ and $b = y^T Y$ such that U and V have the largest possible correlation. Such linear combinations can give insight into the relationships between the two sets of variables. Once x and y are computed, further canonical correlation vectors can be found in the orthogonal directions to the previous ones in the same manner. Standard CCA methods can be found in [2]-[3].

Assume we are given three real matrices $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{n\times n}$, and $C \in \mathbb{R}^{m\times m}$, where $\mathbb{R}$ is the set of real numbers, m, n, p are positive integers such that $p \leq m \leq n$ and B and C are positive definite. The aim of CCA is to find an $n \times p$ transformation x and a $m \times p$ transformation y, such that the matrix $x^T Ay$ is diagonal, $x^T Bx = I$ and $y^T Cy = I$. Thus CCA methods achieve the simultaneous diagonalization of the three matrices. When $B = I$, and $C = I$, we get the familiar singular value decomposition of the matrix A. Here the symbol I denotes an identity matrix of appropriate dimension. Generally, CCA is equivalent to computing the singular value decomposition of the coherence matrix $\mathcal{C} = B^{\frac{-1}{2}} AC^{\frac{-1}{2}}$, where $B^{\frac{-1}{2}}$ and $C^{\frac{-1}{2}}$ are principal square roots of B^{-1} and C^{-1}, respectively.

The three matrices A, B, C involved in CCA are covariance matrices defined as follow. Suppose that $U \in \mathbb{R}^{n\times 1}$ and $V \in \mathbb{R}^{m\times 1}$ are random vectors having means μ_U and μ_V respectively. Then

$$A = E\{(U - \mu_U)(V - \mu_V)^T\}, \quad B = E\{(U - \mu_U)(U - \mu_U)^T\},$$
$$C = E\{(V - \mu_V)(V - \mu_V)^T\},$$

where $E(.)$ denotes the expectation operator. CCA has been generalized in several directions. For example, Leurgans et al.[4] extended CCA to functional data analysis; Kettenring [5] extended two sets CCA to multi-set CCA based on the principle of maximizing some generalized measure of correlation; Luijtens et al.[6] developed linear and nonlinear canonical correlation analysis for group-structured data. In [7] canonical variables used as optimal predictors. Analysis which is based on information theory is given in [8].

In this paper, several dynamical systems which can be seen as a generalization of the singular value decomposition are proposed. Some of these techniques are generalization of Oja's learning systems for principal component analysis, while others use an upper-triangulization process similar to those used in Sanger-type methods [9].

The following notation will be used throughout. The symbols $\mathbb{R}$, and $\mathbb{N}$ denote the set of real numbers, and the set of positive integers, respectively. The derivative of x with respect to time is written as $\dot{x}$. The identity matrix of appropriate dimension is expressed with the symbol I. The derivative of a Lyapunov function $V(x)$ with respect to time along a solution of a dynamical system is denoted by $\dot{V}$. For any square matrix G, the trace of G which is the sum of the diagonal elements of G is denoted bt $tr(G)$. The notation $||x||_2$ denotes the Euclidean norm of x. It will be assumed in this work that all matrices are real.

In the sequel, we endow the space $\mathbb{R}^{n\times p} \times \mathbb{R}^{m\times p}$ with the Riemannian metric $\langle (x_1, y_1), (x_2, y_2)\rangle_R = tr(x_1^T Bx_2 + y_1^T Cy_2)$, for any $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^{n\times p} \times \mathbb{R}^{m\times p}$ and the corresponding norm square $||(x, y||_R^2 = tr(x^T Bx + y^T Cy)$. This dot product defines a positive definite inner product on $\mathbb{R}^{n\times p} \times \mathbb{R}^{m\times p}$ as B and C are positive definite. Recall that for any Riemannian metric $\langle ., .\rangle_R$, an associated gradient vector field $grad\, f(X)$ is defined by the characterizing property $df(X)h = \langle grad\, f(X), h\rangle_R$ for each $X, h \in \mathbb{R}^{n\times p} \times \mathbb{R}^{m\times p}$. Here $f : \mathbb{R}^{n\times p} \times \mathbb{R}^{m\times p} \to \mathbb{R}$ is assumed to be continuously differentiable. Thus, for example, if $f(x, y) = tr(x^T Bx + y^T Cy)$, then $df(x, y)h = \langle grad\, f, h\rangle_R = tr(grad\, f^T diag(B, C)h)$. The gradient associated with the Riemannian metric $\langle ., .\rangle_R$ is seen as $grad\, f(x, y) = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}^{-1} df(x, y)$.

2 Preliminary Results

For completeness, basic concepts from dynamical system theory are summarized in this section. These include Liapunov and Lagrange stability.

2.1 Stability of Dynamical Systems

When a Lyapunov function for a system is known, the direct method is an convenient way of proving stability of equilibria, as Lyapunov's theorem can be used without solving the differential equations. Except for special cases, such as energy functions for mechanical systems, there are no systematic methods to construct Lyapunov functions. Additionally, testing non-negativity of a function is not always an easy task. There are two conditions needed for stability, 1) A positive definite function, 2) The time derivative must be negative definite along any solution of the system. If in addition, 3) the function is radially unbounded, then the system is globally stable.

Let $g(x) : \mathbb{R}^{n\times p} \to \mathbb{R}^{n\times p}$, $p \le n$, be continuously differentiable function and consider the dynamical system

$$\dot{x} = g(x). \tag{1}$$

The point $\bar{x}$ is an equilibrium point for the system (1) if $g(\bar{x}) = 0$. Let $\Omega \subset \mathbb{R}^n$ be a region containing $\bar{x}$ and $V : \Omega \to \mathbb{R}$ be continuously differentiable function such that $V(\bar{x}) = 0$ and $V(x) > 0$ for each $\bar{x} \ne x \in \Omega$, i.e., V is positive definite. Assume also that $\dot{V}(x) \le 0$ for each $x \in \Omega$, i.e., V is negative semi-definite. Then $\bar{x}$ is stable and V is called a Lyapunov function for the system (1) at $\bar{x} \in \Omega$. If $V(x) < 0$ for each $\bar{x} \ne x \in \Omega$, then $\bar{x}$ is asymptotically stable. If in addition to these conditions, we have the function V is radially unbounded, i.e., $V(x) \to \infty$ as $||x|| \to \infty$, then the system is globally stable. The main advantage of using Lyapunov direct method is that Lyapunov theorem can be used to prove stability of equilibria without solving the differential equations. However, constructing Lyapunov functions is not always an easy task. It should be noted that many Lyapunov functions may exist for the same problem. However, a specific choice of Lypunov functions may provide more useful results about the system than others.

Geometrically, the condition $\dot{V} \le 0$ implies that when a trajectory crosses the level surface $V(x) = c$, it moves inside the set $\Omega_2 = \{x \in \mathbb{R}^{n\times p} : V(x) \le c\}$ and remains there. Since V is positive definite, then Ω_2 is bounded and closed, thus the system must converge to some limiting value.

The *domain of attraction* of an equilibrium point $\hat{x}$ of the system (1) is defined as an open set D containing $\hat{x}$ such that for any initial point $x_0 \in D$, the sequence generated by the dynamical system according to (1) with an arbitrarily small step-size $\alpha > 0$ and satisfying $x_k \in D$, for all i) remains in D and ii) x_k converges to $\hat{x}$.

A set S is an *invariant set* for the system (1) if every trajectory $x(t)$ which starts from a point in S remains in S for all time. For example, any equilibrium point is an invariant set. The domain of attraction of an equilibrium point is also an invariant set.

We state next a few stability results for nonlinear autonomous systems. The invariant set theorems reflect the intuition that the decrease of a Liapunov function V has to gradually vanish. In other words $\dot{V}$ has to converge to zero because V is lower bounded.

Theorem 1 (Local Invariant Set Theorem). *Consider an autonomous system of the form $\dot{x} = g(x)$, with g continuous and let $V(x) : \mathbb{R}^n \to \mathbb{R}$ be a scalar function with continuous first partial derivatives. Assume that*

1. *for some $l > 0$, the set Ω_l defined by $V(x) \le l$ is bounded.*
2. *$\dot{V}(x(t)) \le 0$ for all x in Ω_l.*

Let R be the set of all points within Ω_l where $\dot{V}(x) = 0$ and M be the largest invariant set in R. Then, every solution $x(t)$ originating in Ω_l tends to M as $t \to \infty$.

Proof. See Slotine and Li (1991) [10].

In Theorem 1, the word largest means that M is the union of all invariant sets within R. Notice that R is not necessarily connected, nor is the set M. Also, if l in Theorem 1 extends to the whole space $\mathbb{R}^n$, then global asymptotic stability can be established.

Another version of Theorem 1 is stated next.

Theorem 2 (Invariance Principle). *Let Ω be an open set in $\mathbb{R}^n$ that contains an equilibrium point of the system (1). Suppose there exists a function $V : \Omega \to \mathbb{R}$ of class C^1 such that*

1. *V is bounded below, i.e., there is $V_0 \in \mathbb{R}$, such that $V(x) \ge V_0$ for each $x \in \Omega$.*
2. *$\dot{V}(t) \le 0$ along any solution of (1).*
3. *Let E be the set of all points within Ω where $\dot{V}(x) = 0$ and M be the largest invariant set in E, i.e., if $x(t_0) \in \Omega$, where $t_0 \in \mathbb{R}$, then $x(t) \in \Omega$ for $t0 \le t \le t_0 + T$, for some positive $T \in \mathbb{R}$.*

Then, every bounded solution $x(t)$ originating in Ω tends to M as $t \to \infty$.

Proof. The proof of this theorem can be found in [11].

The main difference between Theorems 1 and 2 is that the region Ω_l is dependent on the Liapunov function V, while Ω in Theorem 2 is not dependent on V.

We state next a well known result about Lagrange stability. A dynamical system is Lagrange stable if the continuous state remains bounded from any initial condition. For example, if the continuous state converges to a stationary set, the dynamical system is Lagrange stable.

Theorem 3 (A Lagrange Stability Theorem). *Let W be a bounded neighborhood of the origin and let W^c be its complement (W^c is the set of all points outside W). Assume that $V(x)$ is a scalar function with continuous first partial derivatives in W^c and satisfying:*

1. *$V(x) > 0$ for all $x \in W^c$,*
2. *$\dot{V}(x) \le 0$ for all $x \in W^c$,*
3. *$V(x) \to \infty$ as $||x|| \to \infty$.*

Then each solution of $\dot{x} = g(x)$, is bounded for all $t > 0$.

Proof. The proof of this theorem can be found in [12].

The Lyapunov linearization method explores the relation between the stability of the linearized system with that of the original nonlinear system.

Theorem 4 (Liapunov's Linearization Method). *Let $x = \hat{x}$ be an equilibrium point for the nonlinear system $\dot{x} = g(x)$, where $g : \Omega \to \mathbb{R}^n$ is continuously differentiable and Ω is a neighborhood of $\hat{x}$. Let the Jacobian matrix A at $x = \hat{x}$ be:*

$$A = \frac{\partial g}{\partial x}|_{x=\hat{x}}. \tag{2}$$

Let λ_i, $i = 1, \cdots, n$ be the eigenvalues of A. Then,

1. *The point $\hat{x}$ is asymptotically stable if $Re(\lambda_i) < 0$ for all eigenvalues of A.*
2. *The point $\hat{x}$ is unstable if $Re(\lambda_i) > 0$ for any of the eigenvalues of A.*

Here $Re(\lambda)$ denotes the real part of λ.

Proof. The proof of this theorem can be found in Khalil (2002) [13].

3 CCA Systems Based on Elliptic Constraints

Dynamical systems for computing canonical correlations and variates may be obtained by optimizing the cost function $tr(x^T Ay)$ over the elliptic constraints $x^T Bx = I$ and $y^T Cy = I$, where A, B, C are as defined in the introduction. As shown in [14], the resulting dynamical system is

$$\begin{aligned}\dot{x} &= Ay - Bxy^TA^Tx,\\ \dot{y} &= A^Tx - Cyx^TAy.\end{aligned} \tag{3}$$

Although the system (3) is derived using optimization methods over elliptic regions, simulations have indicated that other variants of this system have similar convergence behavior. These systems are:

$$\begin{aligned}\dot{x} &= Ay - Bxx^TAy,\\ \dot{y} &= A^Tx - Cyx^TAy,\end{aligned} \tag{4a}$$

and

$$\begin{aligned}\dot{x} &= Ay - Bxy^TA^Tx,\\ \dot{y} &= A^Tx - Cyy^TA^Tx.\end{aligned} \tag{4b}$$

To analyze the convergence of the above system define the set

$$\Omega_1 = \{x \in \mathbb{R}^{n\times p}, y \in \mathbb{R}^{m\times p} : x^TAy + y^TA^Tx \text{ is positive definite}\}, \tag{5}$$

where m, n, p are positive integers such that $p \le m \le n$. The behavior of the system (3) is analyzed in the following result.

Theorem 5. *Let $m, n, p \in N$ be positive integers such that $p \le m \le n$ and assume that $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{n\times n}$, and $C \in \mathbb{R}^{m\times m}$ such that B and C are positive definite. Let $x \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^{m\times p}$, and consider the dynamical system (3). Then the system (3) is stable in the sense of Theorem 1 and 2. Then each full rank equilibrium point of this system is stable. Additionally, let $(x(t), y(t))$ be a solution of (3) for $t \ge 0$, and define $\bar{A} = \lim_{t\to\infty} x(t)^TAy(t)$, $\bar{B} = \lim_{t\to\infty} x(t)^TBx(t)$, and $\bar{C} = \lim_{t\to\infty} y(t)^TCy(t)$. If $\bar{A} + \bar{A}^T$ is positive definite, then $\bar{B} = I$, $\bar{C} = I$ and $\bar{A}^T = \bar{A}$.*

Outline of Proof: Stability of equilibrium points follows from Theorem 2 by considering the function V defined by $V(x, y) = \frac{1}{4}tr((x^TBx - I)^2) + \frac{1}{4}tr((y^TCy - I)^2)$. It can be shown that the time derivative of V along the trajectory $(x(t), y(t))$ of the system (3) is

$$\begin{aligned}\dot{V} &= tr\{(x^TBx - I)\{x^TBB^{-1}\{Ay - Bxy^TA^Tx\}\\ &+ (y^TCy - I)\{y^TCC^{-1}\{A^Tx - Cyx^TAy\}\\ &= tr\{(x^TBx - I)\{x^TAy - x^TBxy^TA^Tx\}\\ &+ (y^TCy - I)(y^TA^Tx - y^TCyx^TAy\}\\ &= \frac{1}{2}tr\{(x^TBx - I)\{(x^TAy + y^TA^Tx\\ &- x^TBx(y^TA^Tx + x^TAy)\}\\ &+ \frac{1}{2}(y^TCy - I)\{(x^TAy + y^TA^Tx\\ &- y^TCy(y^TA^Tx + x^TAy)\}\\ &= \frac{1}{2}tr\{(x^TBx - I)\{(x^TAy + y^TA^Tx\}(I - x^TBx)\}\\ &+ \frac{1}{2}(y^TCy - I)\{(x^TAy + y^TA^Tx)(I - y^TCy)\}\\ &\le 0,\end{aligned} \tag{6}$$

for all $(x, y) \in \Omega_1$. Let R be the set of all points within Ω_1 where $\dot{V}(x, y) = 0$, i.e., $(x, y) \in \Omega_2$ where $\Omega_2 = \{x \in \mathbb{R}^{n\times p}, y \in \mathbb{R}^{m\times p} : tr\{(x^TAy + y^TA^Tx)(x^TBx - I)^2 + (y^TCy - I)^2 = 0\}$, and M be the largest invariant set in R. Then, every solution $(x(t), y(t))$ originating in Ω_1 tends to M as $t \to \infty$.

To show that $\bar{B} = I$, $\bar{C} = I$ and $\bar{A}^T = \bar{A}$, let $\bar{B}$, $\bar{C}$, and $\bar{A}$ be as defined above, then as $t \to \infty$, the following equations hold:

$$\bar{A} = \bar{B}\bar{A}^T, \tag{7a}$$

$$\bar{A}^T = \bar{C}\bar{A}. \tag{7b}$$

From these equations it follows that

$$\bar{A} = \bar{B}\bar{A}\bar{B}, \tag{7c}$$

$$\bar{A}^T = \bar{C}\bar{A}^T\bar{C}, \tag{7d}$$

and consequently, for each integer $k \in N$ we have

$$\begin{aligned}\bar{A} &= \bar{B}^k\bar{A}\bar{B}^k,\\ \bar{A}^T &= \bar{C}^k\bar{A}^T\bar{C}^k.\end{aligned}$$

Note that $\bar{B}$ and $\bar{C}$ behave as if they are identity matrices. Indeed it can be proved that $\bar{B} = I$ under the assumption that $\bar{A} + \bar{A}^T$ is positive definite. Let z be an eigenvector of $\bar{B}$ with corresponding eigenvalue λ. This implies that $z^T\bar{A}z = z^T\bar{B}\bar{A}\bar{B}z = \lambda^2 z^T\bar{A}z$. Since $\bar{A} + \bar{A}^T$ and hence $\bar{A}$ is invertible, it follows that $\lambda^2 = 1$, i.e., $\bar{B}^2 = I$, and therefore $\bar{B} = I$ since $\bar{B}$ is positive definite. Similarly, one can prove that $\bar{C} = I$ and consequently $\bar{A}^T = \bar{A}$.

Remark 1: Using Theorem 4, it can be shown that $(x, y) = (0, 0)$ is an unstable equilibrium point for the systems given in (3) and (4). In this case the matrix of the linearized systems at $(0, 0)$ is $\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$ which has positive and negative eigenvalues.

A generalization of Theorem 5 is given in the next result.

Theorem 6. *Let m, n, p, A, B, C, x, and y be as in Theorem 5, and let $K : \mathbb{R}^{n\times p} \times \mathbb{R}^{m\times p} \to: \mathbb{R}^{p\times p}$ be a continuous function such that $x^TAy + y^TA^Tx = \alpha(K + K^T)$ for each x and y, where α is positive number. Consider the following system*

$$\begin{aligned}\dot{x} &= Ay - BxK(x, y),\\ \dot{y} &= A^Tx - CyK(x, y),\end{aligned} \tag{8}$$

then the system is stable with respect to the set Ω_1 (in the sense of Theorem 2).

Outline of Proof: Let Ω_1 be as defined in Theorem 5. We show that if $(x(0), y(0)) \in \Omega_1$, then $(x(t), y(t)) \in \Omega_1$, for $t \ge 0$. By considering a Liapunov function of the form $V(x, y) = \frac{1}{4}tr((x^TBx - \alpha I)^2) + \frac{1}{4}tr((y^TCy - \alpha I)^2)$, it can be shown that the time derivative of V along the trajectory $x(t)$ and $y(t)$ is

$$\begin{aligned}\dot{V} &= tr\{(x^TBx - \alpha I)\{x^TBB^{-1}\{Ay - BxK(x, y)\}\\ &+ (y^TCy - \alpha I)\{y^TCC^{-1}\{A^Tx - CyK(x, y)\}\\ &= tr\{(x^TBx - \alpha I)\{x^TAy - x^TBxK(x, y)\}\\ &+ (y^TCy - \alpha I)(y^TA^Tx - y^TCyK(x, y)\}\\ &= \frac{1}{2}tr\{(x^TBx - \alpha I)\{(x^TAy + y^TA^Tx\\ &- x^TBx(K(x, y) + K(x, y)^T)\}\\ &+ \frac{1}{2}(y^TCy - \alpha I)\{(x^TAy + y^TA^Tx\\ &- y^TCy(K(x, y) + K(x, y)^T)\}\\ &= \frac{1}{2}tr\{(x^TBx - \alpha I)\{(x^TAy + y^TA^Tx\}(\alpha I - x^TBx)\\ &+ \frac{1}{2}(y^TCy - \alpha I)\{(x^TAy + y^TA^Tx)\}(\alpha I - y^TCy)\\ &\le 0,\end{aligned} \tag{9}$$

for all $(x, y) \in \Omega_1$. Note that $\dot{V}$ is computed with respect to the Riemannian metric $\langle ., . \rangle_R$ defined in the introduction. Hence the system is stable in the sense of Theorem 2.

If in Theorem 6, $K(x,y)$ is defined so that $K = \frac{x^TAy+y^TA^Tx}{\alpha}$, then $\frac{\alpha}{2}(K+K^T) = x^TAy + y^TA^Tx$, and therefore we may consider the following system:

$$\begin{aligned} \dot{x} &= Ay - Bx\frac{x^TAy + y^TA^Tx}{2}, \\ \dot{y} &= A^Tx - Cy\frac{x^TAy + y^TA^Tx}{2}, \end{aligned} \tag{10}$$

which corresponds to $\alpha = 2$.

Now assume that $(x(t), y(t))$ is a full rank solution of (10) for $t \geq 0$, then $\lim_{t\to\infty} x(t)^TAy(t)$ is symmetric, $\lim_{t\to\infty} x(t)^TBx(t) = I$, and $\lim_{t\to\infty} y(t)^TCy(t) = I$. To prove this assertion, it will be assumed for convenience that $\alpha = 1$ and thus we have the following system:

$$\begin{aligned} \dot{x} &= Ay - Bx(x^TAy + y^TA^Tx), \\ \dot{y} &= A^Tx - Cy(x^TAy + y^TA^Tx). \end{aligned} \tag{11}$$

As in Theorem 5, let $(x(t), y(t))$ be a full rank solution of (11) for $t \geq 0$ such that $\bar{A} + \bar{A}^T$ is positice definite, where $\bar{A}$, $\bar{B}$, and $\bar{C}$ are as defined above, then

$$\bar{A} = \bar{B}(\bar{A} + \bar{A}^T), \tag{12a}$$

$$\bar{A}^T = \bar{C}(\bar{A} + \bar{A}^T). \tag{12b}$$

Clearly, (12a) and (12b) imply that $\bar{A}\bar{B}$ and $\bar{A}^T\bar{C}$ are symmetric and therefore the following equation holds

$$(\bar{A} + \bar{A}^T)\bar{A} = \bar{A}^T(\bar{A} + \bar{A}^T). \tag{13a}$$

Equation (13a) implies that $\bar{A}^2$ is symmetric, i.e.,

$$\bar{A}^2 = (\bar{A}^T)^2. \tag{13b}$$

Theorem 9 (see Appendix) guarantees that $\bar{A}^T = \bar{A}$. Additionally, Equations (12a) and (12b) yield $\bar{B} = \frac{1}{2}I$ and $\bar{C} = \frac{1}{2}I$.

3.1 Sanger's Type Learning Systems

As shown in Theorem 5, $x(t)$ and $y(t)$ do not converge to the actual canonical variates since $\bar{A}$ is generally not diagonal. The true canonical variates can be recovered by incorporating an upper-triangulization in the system (3), similar to that of [9], as follows:

Theorem 7. *Let m, n, p, A, B, C, x, and y be as in Theorem 5, and consider the following system:*

$$\begin{aligned} \dot{x} &= Ay - Bx\{UT(x^TAy) + LT(x^TAy) - dd(x^TAy)\}, \\ \dot{y} &= A^Tx - Cy\{UT(x^TAy) + LT(x^TAy) - dd(x^TAy)\}. \end{aligned} \tag{14}$$

Then each full rank equilibrium point of this system is asymptotically globally stable over the set Ω_1 defined in (5). Additionally, if $(x(t), y(t))$ is a full rank solution of (6) for $t \geq 0$, then $\lim_{t\to\infty} x(t)^TAy(t) = \Sigma$, $\lim_{t\to\infty} x(t)^TBx(t) = I$, and $\lim_{t\to\infty} y(t)^TCy(t) = I$. If $\bar{A} + \bar{A}^T$ is positive definite, then $\bar{B} = I$, $\bar{C} = I$ and $\bar{A}$ is diagonal. Here Σ is a diagonal matrix whose diagonal elements are the canonical correlations of the triplet (A, B, C). The functions $UT(.)$ $(LT(.))$ set the entries of the lower (upper) triangular part of (.) to zeros, and keeps the main diagonal, and the upper(lower)-triangular part of (.) unchanged. The matrix $dd(.)$ is diagonal whose diagonal elements are the diagonal entries of (.).

Outline of Proof: Stability of equilibrium points of the system (14) follows from Theorem 2 and Theorem 7 as follows. Let $V(x,y) = \frac{1}{4}tr((x^TBx - I)^2 + (y^TCy - I)^2)$, then $V(x,y)$ is positive semi-definite and the time derivative of V along any trajectory of (17) is

$$\begin{aligned} &\dot{V}(x,y) = tr\{(x^TBx - I)(x^TAy + (x^TAy)^T \\ &- x^TBx(UT(x^TAy) + UT(x^TAy)^T \\ &+ LT(x^TAy) + UT(x^TAy)^T - 2dd(x^TAy)) \\ &+ (y^TCy - I)(y^TA^Tx + x^TAy - y^TCy(x^TAy + (x^TAy)^T \\ &- x^TBx(UT(x^TAy) + UT(x^TAy)^T + LT(x^TAy) \\ &+ UT(x^TAy)^T - 2dd(x^TAy))\} \\ &= -tr(\{(x^TBx - I)^2 + (y^TCy - I)^2)\}\{x^TAy + (x^TAy)^T\}) \\ &\leq 0, \end{aligned}$$

for each $x \in \Omega_1$. The time derivative $\dot{V}(x,y)$ is computed with respect to the Riemannian metric $\langle .,. \rangle_R$ defined in the introduction. Hence the system (14) is stable over the set Ω_1.

Now let $\bar{A} = U + D + L$, where U and L are upper and lower triangular matrices, respectively and D is diagonal. Note that the diagonal elements of U and L are zeros. We show that $\bar{A}$ is diagonal under the assumption that $\bar{A} + \bar{A}^T$ is positive definite. As $t \to \infty$, it follows from (14) that

$$\bar{A} = \bar{B}(U + D + L^T), \tag{15a}$$

$$\bar{A}^T = \bar{C}(U + D + L^T). \tag{15b}$$

Since $\bar{B}$ and $\bar{C}$ are symmetric, then $(U + D + L^T)^T\bar{A}$ and $(U + D + L^T)^T\bar{A}^T$ are also symmetric:

$$(U + D + L^T)^T\bar{A} = \bar{A}^T(U + D + L^T). \tag{15c}$$

$$(U + D + L^T)^T\bar{A}^T = \bar{A}(U + D + L^T). \tag{15d}$$

By adding the equations (15c) and (15d) and observing that $\bar{A} + \bar{A}^T = U + L + D + U^T + L^T + D$ we obtain the following equations:

$$(U + D + L^T)^T(\bar{A} + \bar{A}^T) = (\bar{A} + \bar{A}^T)(U + D + L^T),$$

or equivalently,

$$(U + D + L^T)^{2T} = (U + D + L^T)^2.$$

It follows from Theorem 9 (see Appendix) that the last equation holds only if $(U + D + L^T)^T = U + D + L^T$, or $(U + L^T)^T = U + L^T$. Therefore, $(U + L^T)^T = U + L^T = 0$.

$$U = -L^T = 0$$

Hence $U = -L^T = 0$, and $\bar{A} = D$. This also shows that $\bar{B} = I$ and $\bar{C} = I$.

Remark 2: A version of Theorem 7 applied to the system (11) yields the following CCA system:

$$\begin{aligned} \dot{x} &= Ay - BxUT(x^TAy + y^TA^Tx), \\ \dot{y} &= A^Tx - CyUT(x^TAy + y^TA^Tx). \end{aligned} \tag{16}$$

Let $(x(t), y(t))$ be a full rank solution of (14) for $t \geq 0$, and let $\bar{A}$, $\bar{B}$, and $\bar{C}$ be as defined above. Assume that $\bar{A} + \bar{A}^T = U + L$, where U and L are upper and lower triangular matrices, respectively and the diagonal elements of L are zeros. Then

$$\bar{A} = \bar{B}U \text{ and } \bar{A}^T = \bar{C}U. \tag{17}$$

Since $\bar{B}$, and $\bar{C}$ are symmetric, it follows from (17) that

$$U^T\bar{A} = \bar{A}^TU \text{ and } U^T\bar{A}^T = \bar{A}U. \tag{18}$$

By adding the equations in (18) we obtain

$$U^T(\bar{A}^T + \bar{A}) = (\bar{A}^T + \bar{A})U,$$

or

$$U^T(U + L) = (U^T + L^T)U,$$

from which it follows that

$$U^T L = L^T U.$$

If it is assumed that $\bar{A}+\bar{A}^T$ is positive definite, then $L = 0$, i.e., $\bar{A} + \bar{A}^T = U$. Therefore, since $\bar{A} + \bar{A}^T$ is symmetric, we have

$$\bar{A} + \bar{A}^T = D,$$

for some diagonal matrix D. It remains to show that $\bar{A}$ is diagonal. From Equation (18) it follows that $U^T D = DU$, or equivalently, $(L^T + D)D = D(D + L)$. Hence $L^T D = DL$ which implies that $L = 0$ since D is positive definite. This proves that $\bar{A} = \frac{D}{2}$, and hence $\bar{B} = \frac{1}{2}I$, $\bar{C} = \frac{1}{2}I$. Although, the system (16) may not be a gradient system, numerical simulations not included here, have indicated that $x(t)^T Ay(t)$ converges to a diagonal matrix whose trace maximizes $trace(x^T Ay)$ over the constraints $x^T Bx = I$, and $y^T Cy = I$.

4 Systems Based on Logarithmic Cost Function

Faster convergent CCA dynamical systems may be derived from optimizing the unconstrained cost function $F(x, y) = \frac{1}{2}tr\{\log(x^T Ay) - x^T Bx - y^T Cy\}$ over the set Ω_1 defined in (5). Note that the natural logarithm of a square matrix Z, denoted by $\log(Z)$, is defined if and only if Z is invertible. This means that $\log(Z)$ is defined as long as the spectrum of Z does not contain the origin.

Thus the gradient flow corresponding to the cost function $F(x, y)$ is

$$\begin{aligned} \dot{x} &= Ay(x^T Ay)^{-1} - Bx, \\ \dot{y} &= A^T x(y^T A^T x)^{-1} - Cy. \end{aligned} \tag{19}$$

It should be noted that $(x^T Ay)^{-1}$ always exists as long as $(x, y) \in \Omega_1$. It is immediately clear from Theorem 3 that by utilizing the function $V(x, y) = tr(x^T x + y^T y)$, System (19) is Lagrange stable, i.e., each solution of (19) is bounded for $t \geq 0$. Further convergence analysis of (19) may be established via the Liapunov function $V(x, y)$ defined over Ω_1 as $V(x, y) = \frac{1}{4}tr((x^T Bx - I)^2 + (y^T Cy - I)^2)$. Then $V(x, y)$ is positive semi-definite (and thus lower bounded) and the time derivative of V along any trajectory of (19) is

$$\begin{aligned} &\dot{V}(x, y) = tr\{(x^T Bx - I)(x^T Ay(x^T Ay)^{-1} \\ &+ y^T A^T x(y^T A^T x)^{-1} - x^T Bx\} \\ &+ (y^T Cy - I)(y^T A^T x(y^T A^T x)^{-1} + x^T Ay(x^T Ay)^{-1} - y^T Cy\} \\ &= -tr((x^T Bx - I)^2 - (y^T Cy - I)^2) \leq 0, \end{aligned}$$

for each $(x, y) \in \Omega_1$. The time derivative $\dot{V}(x, y)$ is computed with respect to the Riemannian metric $\langle ., . \rangle_R$ defined in the introduction. Hence the system (19) is stable over the set Ω_1. It should be noted that the set Ω_1 may be replaced with the set $\Omega_3 = \{(x, y) : x^T Ay \text{ is invertible}\}$.

Remark 3: Numerical simulations have indicated that different versions of (19), which are of Sanger's type, have similar convergent behavior. Two of these variants are given below:

$$\begin{aligned} \dot{x} &= AyUT((x^T Ay)^{-1}) - Bx, \\ \dot{y} &= A^T xUT((x^T Ay)^{-1}) - Cy, \end{aligned} \tag{20}$$

and

$$\begin{aligned} \dot{x} &= Ay(UT(x^T Ay))^{-1} - Bx, \\ \dot{y} &= A^T x(UT(x^T Ay))^{-1} - Cy. \end{aligned} \tag{21}$$

4.1 Convergence Properties of (20) and (21)

Let $(x(t), y(t))$ be a full rank solution of (20) for $t \geq 0$, and assume that $(x(0), y(0)) \in \Omega$. Let $\bar{A}$, $\bar{B}$, and $\bar{C}$ be as defined above. Assume that $\bar{A}^{-1} = U + L$, where U and L are upper and lower diagonal matrices, respectively and the diagonal elements of L are zeros. As $t \to \infty$, equation (20) yields

$$\bar{A}U = \bar{B}, \tag{22a}$$

$$\bar{A}^T U^T = \bar{C}. \tag{22b}$$

Substituting $U = \bar{A} - L$ in Equation (20a) gives

$$\bar{A}(\bar{A}^{-1} - L) = \bar{B},$$

or equivalently, $\bar{A}L = I - \bar{B}$ which is symmetric. Consequently, the following equations hold:

$$\bar{A}L = L^T \bar{A}^T,$$

$$\bar{A}^{-1} L^T = L\bar{A}^{-T},$$

and

$$(U + L)L^T = L(U^T + L^T).$$

It follows from these equations that

$$UL^T = LU^T.$$

Since $\bar{A}$ is assumed to be positive definite, then

$$L = 0, \text{ and } \bar{A} = U.$$

From Equation (22a), $U^2 = \bar{B}$ is symmetric, i.e., $(U^T)^2 = U^2$. Theorem 9 guarantees that $U^T = U = D$. Note that only one equation of (22) is sufficient to prove that $\bar{A}$ is diagonal.

Similarly, to show that $\bar{A}$, $\bar{B}$, and $\bar{C}$ obtained in System (21) are diagonal, assume that $\bar{A} = U + L$, where U and L are upper and lower diagonal matrices, respectively and the diagonal elements of L are zeros. Hence,

$$\bar{A}U^{-1} = \bar{B}, \tag{23a}$$

and

$$\bar{A}^T U^{-1} = \bar{C}. \tag{23b}$$

Since the matrices $\bar{B}$ and $\bar{C}$ are symmetric, then

$$\bar{A}U^{-1} = U^{-T}\bar{A}^T,$$

or equivalently,

$$\bar{A}^T U = U^T \bar{A},$$

which imply that $(U^T + L^T)U = U^T(U + L)$, and hence $L^T U = U^T L$. Since it is assumed that $\bar{A} + \bar{A}^T$ is positive definite, then

$$L = 0, \text{ and } \bar{A} = U.$$

Also, from Equation (23b), we obtain the following relations:

$$\begin{aligned} \bar{A}^T U^{-1} &= U^{-T}\bar{A}, \\ U^T U^{-1} &= \bar{C}, \\ U^T U^{-1} &= U^{-T} U, \\ (U^T)^2 &= U^2. \end{aligned}$$

Applying Theorem 9, it follows that $U^T = U = D$, and hence $\bar{B} = I$ and $\bar{C} = I$.

Remark 4: Other modifications of the dynamical systems (20) and (21) can be considered. Some of these systems are given in the equations (24) and (25) below:

$$\begin{aligned} \dot{x} &= AyUT((x^T Ay)^{-1}) - Bx, \\ \dot{y} &= A^T xUT((y^T A^T x)^{-1}) - Cy, \end{aligned} \tag{24}$$

and

$$\begin{aligned} \dot{x} &= Ay(UT(x^T Ay))^{-1} - Bx, \\ \dot{y} &= A^T x(UT(y^T A^T x))^{-1} - Cy. \end{aligned} \tag{25}$$

The convergence behavior of Systems (24) and (25) may be shown as follows. Let $x(t), y(t)$, $\bar{A}, \bar{B}, \bar{C}$ be as above and assume that

$$\bar{A} = U + D + L, \tag{26a}$$

then the system (24) implies that

$$\bar{A}(U + D)^{-1} = \bar{B}, \tag{26b}$$

and

$$\bar{A}^T(D^T + L^T)^{-1} = \bar{C}. \tag{26c}$$

A few algebraic manipulations of the equations (26a)-(26c) yield the following relations:

$$\bar{A}(\bar{A} - L)^{-1} = \bar{B} = (\bar{A} - L)^{-T}\bar{A}^T, \tag{27a}$$

$$(\bar{A} - L)^T\bar{A} = \bar{A}^T(\bar{A} - L), \tag{27b}$$

$$L^T\bar{A} = \bar{A}^T L. \tag{27c}$$

By incorporating (26a) into (27c), wo obtain

$$L^T(L + D + U) = (L^T + D + U^T)L, \tag{28a}$$

and

$$L^T(D + U) = (D + U^T)L. \tag{28b}$$

Hence $L = 0$ provided that $D + U + D + U^T = 2D + U + U^T$ is positive definite. This implies that

$$\bar{A} = D + U,$$

and consequently,

$$(U^T + D)D^{-1} = D^{-1}(U + D),$$

$$U^T D^{-1} = D^{-1}U.$$

This proves that $U = 0$ and hence $\bar{A} = D$, provided that $\bar{A}+\bar{A}^T$ is positive definite.

Similar proof can be applied to the system (25) to show that $\bar{A}, \bar{B}, \bar{C}$ are diagonal. Although the above proof shows that $\bar{A}$ is diagonal, simulations have shown that Iterations (24) and (25) are very slow.

5 Conclusion

Canonical variate dynamical systems have been derived and analyzed in this paper. Some of these systems were obtained by utilizing optimization techniques with elliptic constraints, while others are resulted from unconstrained optimization of appropriate cost functions. One may view the proposed flows as generalization of Oja's and Sanger's principal component flows to canonical correlation analysis. It should be pointed out that only proof outlines are given for some of the results of this work. More rigorous proofs are under consideration. Furthermore, numerical stability and convergence need to be explored especially when dealing with complex data and matrices and also in relation to non full rank initial conditions.

6 Appendix

In this section, we state some results from matrix theory. These results provide conditions under which certain matrices are symmetric or diagonal.

Theorem 8. *Let $A, D_1, D_2, D_3 \in \mathbb{R}^{n\times n}$ and assume that $d_i + d'_j \neq 0$ for each $i \neq j$, where d_i and d'_j are eigenvalues of D_1 and D_2, respectively. If $D_1A + AD_2 = D_3$, then A is diagonal. Specifically, this result holds true if D_1 and D_2 are both positive definite or both negative definite diagonal matrices.*

Proof: Assume that $A = [a_{ij}]$, $D_1 = \mathrm{diag}(d_1, \cdots, d_n)$, and $D_2 = \mathrm{diag}(d'_1, \cdots, d'_n)$. For each $i \neq j$ we have $a_{ij}d_j + d'_i a_{ij} = 0$ or $(d'_j + d_i)a_{ij} = 0$. Since $d'_j + d_i \neq 0$ by assumption, then $a_{ij} = 0$ for $i \neq j$, i.e., A is diagonal.

Theorem 9. *Let $A \in \mathbb{R}^{n\times n}$ and assume that all eigenvalues of $A + A^T$ are positive. If $A^2 = (A^T)^2$, then $A^T = A$.*

Proof: Let $F = A + A^T$, then

$$A^2 = (A^T)^2 = (F - A)^2,$$

$$A^2 = (F - A)^2 = F^2 - AF - FA + A^2,$$

$$AF + FA = F^2.$$

Let $F = U\Sigma U^T$ be an eigendecomposition of F, where Σ is diagonal, and U is orthogonal. Then

$$AU\Sigma U^T + U\Sigma U^T A = U\Sigma^2 U^T.$$

Pre- and post-multiplying the last equation with U^T and U yield:

$$U^T AU\Sigma + \Sigma U^T AU = \Sigma^2.$$

It follows from Theorem 8 that

$$U^T AU = \Sigma_1$$

for some diagonal matrix Σ_1. Hence A can be expressed as

$$A = U\Sigma_1 U^T = A^T.$$

References

[1] L. M. Ewerbring and F. T. Luk, "Canonical correlations and generalized SVD: Applications and new algorithms," J. Comput. Applied Math., 27 (1989), pp. 37-52.

[2] T.W. Anderson, Introduction to Multivariate Statistical Analysis, Wiley, New York,1958.

[3] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis,Academic Press, New York, 1979.

[4] S.E. Leurgans, R.A. Moyeed, B.W. Silverman, "Canonical correlation analysis when the data are curves," J. Roy.Statist.Soc.Ser.B 55 (3)(1993)725-740.

[5] J.R. Kettenring, "Canonical analysis of several sets of variables," Biometrika 58 (1971), 433 451.

[6] K. Luijtens, F. Symons, M. Vuylsteke-wauters, Linear and non-linear canonical correlation analysis: an exploratory tool for the analysis of group-structured data, J.Appl.Statist.21 (3)(1994)43 61.

[7] V. J. Yohai; M. S. Garcia Ben, "Canonical Variables as Optimal Predictors," The Annals of Statistics, Vol. 8, No. 4. (Jul., 1980), pp. 865-869.

[8] Xiangrong Yin, "Canonical correlation analysis based on information theory," Journal of Multivariate Analysis 91 (2004)161 176.

[9] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward network," Neural Networks, 2:459-473, 1989.

[10] J. J. E. Slotine and W. Li. Applied nonlinear Comtrol. Prentice Hall, 1991.

[11] R. Balan, "An Extension of Barbashin-Krasovski-LaSalle Theorem to a Class of Nonautonomous Systems," technical report, Princeton University, 1995.

[12] J. LaSalle, "Some Extensions of Liapunov's Second Method," IRE Transactions on Circuit Theory, Volume 7,Issue 4,Dec 1960 Page(s):520-527.

[13] Hassan K. Khalil, Nonlinear Systems, 3rd edition. Prentice Hall, 2002.

[14] M.A. Hasan, "Lagrangian Gradient for Principal Singular Component Analysis," IEEE International Symposium on Circuits and Systems, 2007. ISCAS 2007. 27-30 May 2007, Page(s):2315-2318.